Big Data Challenges faced by Organizations

Noureen Kausar
[1]Department of Information Technology, GC University
Faisalabad, Pakistan.
nooreallah@hotmail.com

Rabia Saleem
Department of Information Technology, GC University
Faisalabad, Pakistan.
rabia_278@hotmail.com

Sidra Amin
Department of Information Technology, GC University
Faisalabad, Pakistan.
Sidra_chudry44@yahoo.com

***Abstract* — Big data is a term that describes a large or complex data volume. That data volume can be processes using traditional data processing software or techniques that are insufficient to deal with them. But big data is often noisy, heterogeneous, irrelevant and untrustworthy. As the speed of information growth exceeds Moore's Law at the beginning of this new century, excessive data is making great troubles to human beings. However this data with special attributes can't be managed and processed by the current traditional software system, which become a real problem. In this paper was discussed some big data challenges and problems that are faced by organizations. These challenges may relate heterogeneity, scale, timelines, privacy and human collaboration. Survey method was used as a theoretical solution framework. Survey method consists of a questionnaires report. Questionnaires report consists of all challenges and problems faced by organizations. After knowing the problem and challenges of organizations, a solution was given to organization to solve big data challenges.**

Keywords: Big data, Heterogeneity, Human Collaboration, Organizations Problems, challenges, Security

I. INTRODUCTION

"Huge records are like teenage sex: absolutely everyone talks about it. No one without a doubt knows a way to do it. Everyone thinks everybody else is doing it. So, all people claim they're doing it, too." The concept of large information has been endemic inside computer science for the reason those earliest days of computing. "Massive information" at the start intended the quantity of facts that could not be processed (successfully) through traditional database techniques and gear. Every time a new garage medium was invented, the amount of records reachable exploded because it is able to be effortlessly accessed. The explosion of statistics has not been followed with the aid of a corresponding new garage medium [1, 21, 22].

We outline "large statistics" as the quantity of facts just past technology's capability to keep, manage and procedure. These imitations are best found by means of a robust analysis of the information itself, express processing needs and the capabilities of the tools (hardware, software, and strategies) used to research it. As with every new trouble, the realization of how to continue may additionally result in an advice that new tools want to be cast to carry out the new duties. As little as five years in the past, we have been only deliberating tens to loads of gigabytes of storage for our non-public computers. Today, We're wondering in tens to masses of terabytes. As a consequence, big records are a shifting goal. placed some other manner, it's far that quantity of records that is simply past our instant draw close, e.g., we should paintings tough to shop it, get right of entry to it, manage it, and technique it [2,24]. In august 2010, the white residence, OMB, and ostp proclaimed that huge facts are a national mission and precedence at the side of healthcare and national protection (aip, 2010). the country wide technological know-how foundation, the countrywide institutes of fitness, the u.s. geological survey, the departments of defense and power, and the defense superior research initiatives corporation announced a joint r&d initiative in march 2012 with the intention to make investments greater than $2 hundred million to increase new big records tools and techniques. Its purpose is to enhance our "…know-how of the technologies had to manipulate and mine big amounts of records; observe that understanding to other medical fields "in addition to cope with the countrywide dreams inside the areas of health power protection, education and researcher [3, 27].

A. Big Data has changed the way

Massive statistics has changed the way that we undertake in doing groups, managements and researches. Statistics-in depth technology especially in statistics-in depth computing is coming into the arena those goals to offer the gear that we need to handle the huge records troubles [4, 25, 26] Facts-extensive science is emerging as the fourth clinical paradigm in phrases of the previous specifically empirical technology, theoretical technological know-how and computational technological know-how. Thousand years in the past, scientists describing the herbal phenomenon only primarily based on human empirical evidences, so we call the science at that point as empirical science [5].

B. Relational database management systems

Relational database management systems and computer facts- and visualization-packages frequently have trouble

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

managing huge data. The paintings might also require "hugely parallel software program walking on tens, loads, or maybe heaps of servers". What counts as "massive statistics" varies depending on the competencies of the users and their tools, and expanding abilities make huge data a shifting target. "For some organizations, dealing with loads of gigabytes of facts for the primary time may additionally cause a need to rethink facts management alternatives. For others, it could take tens or loads of terabytes before facts size turns into a huge attention" [6].

C. *Big Data: What's All the Fuss About?*

"Each days, we create as a good deal statistics as we did from the dawn of civilization up until 2003" Eric Schmidt, former Google ceo, the belief of massive records is not completely new. In the end, cfos are conversant in managing mounting volumes of information. So why all of the fuss? The volumes in maximum middle financial packages are big but clearly now not inside the nation-states of the terabytes, petabytes, or even zetta bytes being generated by means of the billions of linked gadgets purchasers, companies, and governments use every day round the arena. Large statistics takes 'large' to an entirely new stage, not just in the quantity of records available, but inside the monetary possibilities that information can generate.

But there are different motives that make the troubles of large records one of a kind and urgent. First, the distance among the opportunities afforded with the aid of large records and an organization's capability to take advantage of it's far widening via the second one. As an example, statistics is expected to grow globally with the aid of forty percent in line with yr. however increase in it spending is languishing at simply 5 percent.

2d, companies are being ravaged concurrently by means of the twin demanding situations of rampant financial, regulatory and marketplace alternate and unheard of volatility - all going on at close to 'twitter-speed'. As an end result, there may be excessive hobby in technologies and strategies that can provide a side and shine a mild on market developments fast beforehand of competitors.

Third, stirred by way of large information successes reported inside the retail, healthcare and financial services sectors, amongst others, a few marketplace observers do not forget we are at the point of inflection, i.e. that massive records truly is the catalyst for absolutely new boom possibilities, products and services in the personal area, now not to mention price savings and more effective useful resource allocation for government groups [7].

D. *Big Data Challenges by Alexandru*

Monetary entities and no longer simplest, had advanced over time new and greater complicated strategies that allows them to look marketplace evolution, their function on the market, the efficiency of supplying their services and/or merchandise and so forth. For being able to accomplish that, a large quantity of records is wanted in order to be mined so that could generate treasured insights. Every yr. the facts transmitted over the net is developing exponentially. By means of the give up of 2016, cisco estimates that the yearly international statistics site visitors will reach 6.6 zettabytes. The task might be not simplest to "accelerate" the net connections, but also to expand software systems with a view to be capable of deal with big data requests in most effective time. To have a higher understanding of what big information means, the table below represents a comparison among conventional statistics and large facts (know-how big records) [8].

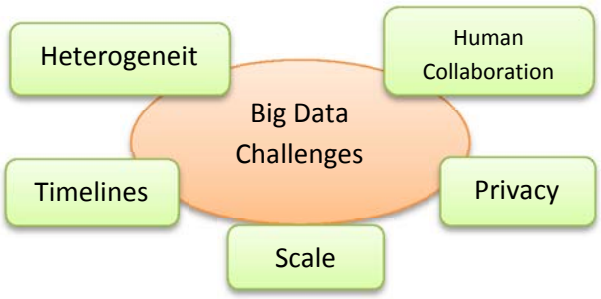
TABLE I. Big Data By Alex

Understanding Big Data	
Traditional Data	**Big Data**
Documents Finances Stock Records Personnel files	Photos Audio and Video 3D Models Simulations Location data

This situation gives data about the quantity and the sort of huge statistics. It is difficult to paintings with complicated statistics on trendy database structures or on personal computer systems. Generally it takes parallel software program systems and infrastructure that may manage the process of sorting the quantity of statistics that, for instance, meteorologists want to analyze. the request for extra complicated records is getting higher each yr. streaming data in actual-time is turning into a challenge that ought to be triumph over via those corporations that gives such services, as a way to hold their role on the market. Via collecting records in a digital form, corporations take their improvement to a brand new degree. Analyzing virtual data can speed the method of making plans and can also display styles that may be further used so one can improve techniques. Receiving statistics in real-time about consumer needs is useful for seeing market trends and forecasting.

II. BIG DATA CHALLENGES FACED BY ORGANIZATIONS

Big data challenges that is discussed in my research, has been shown in Figure 1.



https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Figure1. Big Data problems

A. Heterogeneity and Incompleteness

Whilst people consume records, a high-quality deal of heterogeneity is without problems tolerated. In truth, the nuance and richness of natural language can offer precious depth. However, gadget analysis algorithms assume homogeneous data, and cannot understand nuance. In consequence, facts have to be carefully established as a primary step in (or previous to) records evaluation. Remember, as an instance, a patient who has multiple medical methods at a medical institution [9].

B. Scale

Of path, the first component everybody thinks of with huge statistics is its length. In spite of everything, the word "big" is there within the very call. Managing large and hastily increasing volumes of facts has been a difficult trouble for many decades. Inside the past, this challenge was mitigated through processors getting quicker, following Moore's regulation, to offer us with the assets needed to address growing volumes of statistics. But there is an essential shift underway now: information quantity is scaling quicker than compute resources, and CPU speeds are static [10].

C. Timeliness

The turn facet of length is speed. The larger the information set to be processed, the longer it'll take to investigate. The layout of a machine that efficiently deals with length is probable also to result in a device that may system a given size of data set faster [11].

D. Privacy

The privateers of facts are another huge concern, and one that increases within the context of big statistics. For digital fitness facts, there are strict legal guidelines governing what can and can't be executed. For different records, regulations, in particular in the us, are less forceful. However, there may be excellent public fear regarding the beside the point use of private information, mainly via linking of statistics from multiple resources. Dealing with privacy is effectively each a technical and a sociological problem, which should be addressed collectively from each views to comprehend the promise of large information [12, 23].

E. Human Collaboration

No matter the first rate advances made in computational evaluation, there remain many patterns that humans can effortlessly detect but laptop algorithms have a hard time finding. Certainly, catches take advantage of exactly this truth to tell human internet users aside from computer programs. Ideally, analytics for big data will no longer be all computational – rather it will likely be designed explicitly to have a human within the loop. The new sub-field of visible analytics is making an attempt to do that, at the least with admire to the modeling and evaluation phase inside the pipeline. There's similar fee to human input at all stages of the analysis pipeline [13].

The organizations that are includes for survey of big data challenges faced, these are following

- Government College Lahore
- UVAS
- Punjab University
- UET
- GCUF
- University of Agriculture, Faisalabad
- Faisalabad Institute of Cardiology
- FESCO
- Mobilink and Warid Company
- U Phone Company
- Zong Company
- Wateen Telecom

All these organizations have some related problems, but some educational institutions are still not using big data tools for saving data, only some organization using big data tools for saving data.

These are following related problems faced by organizations during saving data

- Eliminate data entry errors
- Test survey designs
- Change mind
- Try to get better result
- Developed in house
- Space
- Missing data
- Redundancy
- Data collection process
- Human collaboration
- Online verifying
- Information missing
- Incomplete data
- Empty source file
- Saving data
- Security

III. MATERIAL AND METHODS

I was survey of big data challenges in twelve organizations. These are all organizations categorized into four broad area .i.s.

- Educational Big Data Challenges
- Big Data Challenges in Telecommunication System
- Big Data Challenges in Hospital
- Big Data Challenges in Electrical Power System

These are all organization has been shown in Figure 2.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

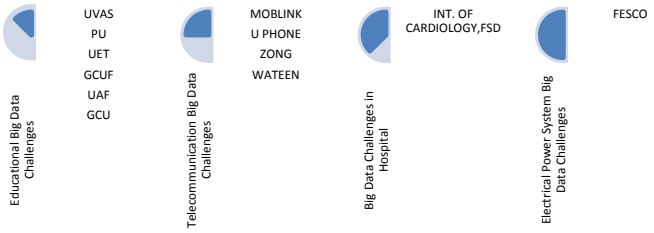


Figure2. Illustrate the survey report about organizations in Pakistan

A. *Educational Big Data Challenges*

Establishments of better education are running in an increasingly more complex and competitive environment. they may be under increasing pressure to respond to national and global financial, political and social change such as the growing want to boom the percentage of students in sure disciplines, embedding place of business graduate attributes and making sure that the exceptional of gaining knowledge of applications are both nationally and globally applicable [14].

I was survey of many educational institutions of Pakistan. During my survey I realized, educational institutions have faced many problems just because, they are not using Big Data tools for saving their data. All the problems have been shown in Figure 3.

Figure3. Illustrate Big Data related problems in Educational System of Pakistan

B. *Big Data Challenges in Telecommunication System*

Within the era of Telecommunication, nearly every huge enterprise encounters big information issues, mainly for multinational agencies [15]. On the only hand, the ones corporations commonly have a big variety of customers around the arena. Alternatively, there are very huge volume and speed of their transaction records. For instance, FICO's falcon credit score card fraud detection system manages over 2.1 billion legitimate debts round the arena. There are above three billion portions of content material generated on Facebook every day. The same problem occurs in each internet agencies. The list may want to pass on and on, as we witness the future agencies warfare fields focusing on large facts [16].

The future of telecommunication has been shown in Figure 4.

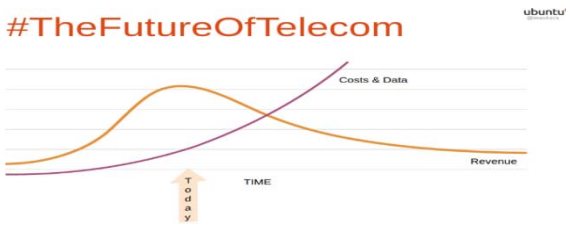


Figure4. Illustrate Future of telecom

C. *Big Data Challenges in Hospital*

The health network is facing a tsunami of health- and healthcare-associated content generated from several affected person care points of contact, state-of-the-art scientific instruments, and web-primarily based health communities [17, 18].

I was survey of Hospital of Pakistan. During my survey I realized, they have also faced many problems just because, they are not using Big Data tools for saving their data. All the problems have been shown in Figure 5.

Figure5. Illustrate Big Data related problems in Hospital of Pakistan

D. *Big Data Challenges in Electrical Power System*

A strength grid is a complicated system connecting an expansion of electrical electricity mills to customers via strength transmission and distribution networks across a massive geographical place, as illustrated in determine 1 [19]. The safety and reliability of energy grids has crucial effect on society and people's each day lifestyles. As an example, on August 14, 2003, a huge part of the Midwest and Northeast United States and Ontario, Canada, skilled an electric powered strength blackout, which affected an area with a population of about 50 million humans. The envisioned overall prices range between $four billion and $10 billion (U.S. greenbacks) inside the use, and $2.3 billion (Canadian dollars) in Canada [20].

I was also survey of FESCO that is an institute of electrical power system of Pakistan. During my survey I realized, they

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

have also faced many problems just because, they are not using Big Data tools for saving their data. All the problems have been shown in Figure 6.

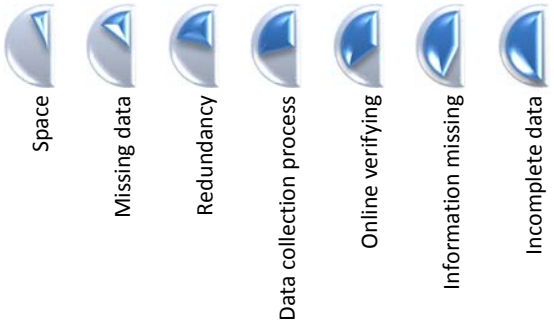


Figure6. Illustrate Big Data related problems in Electrical Power System

IV. RESULTS AND DISCUSSION

Survey research is one of the maximum essential regions of dimension in carried out social research. The wide location of survey research encompasses any size procedures that involve asking questions of respondents. A "survey" may be something forms a short paper-and-pencil feedback shape to an extensive one-on-one in-depth interview. Survey research is categorized into two broad types' i.s interview and questionnaire. Questionnaire report is consist total 30 questions. I was survey total twelve organizations in Pakistan and filled questionnaire from all these organizations.

Questionnaire were included manage raw data, strategies used for saving data, data saved for future use, Challenges faced during data collection and saving, big data tools are used for saving data, generating source used for data recording, format used for information extracting and changing, method adopt for data cleaning, methods used for querying data, tools used for mining data, interpreted result, type of error you face while managing your data, backup of data, database system used for saving data, power source to use for always on system, type of application, types of database are used for manage your data, type of model, client-server based and type of locking etc.

Organizations were included Moblink, U phone, Zong, Wateen, Cardiology hospital, FSD, GCUF, GCU, PU, UET, UVAS, UAF and FESCO etc.

A. Managing of Raw Data

In this table illustrate that 75% organizations of Pakistan are managed their raw data using computer system, and 25% organizations managed their raw data using both computer and manual system. The results also have been shown in Table 2 and in Figure 7.

TABLE II. Raw Data Management

Manage Raw Data

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Computerized	9	75.0	75.0	75.0
	Both	3	25.0	25.0	100.0
	Total	12	100.0	100.0	

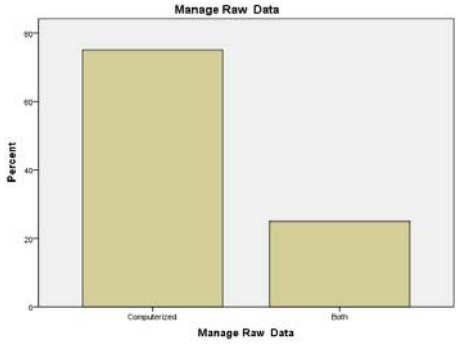


Figure7. Illustrate management of raw data

B. Strategies for Saving Data

In this table shows that 58.3% organizations of Pakistan using backup system, 8.3% using cloud computing, 8.3% using data warehouses, 16.7 using no strategies and 8.3% using others strategies for saving their data. The results also have been shown in Table 3 and in Figure 8.

TABLE III. Strategies For Saving Data

Strategies For Saving Data

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Backup system	7	58.3	58.3	58.3
	Cloud computing	1	8.3	8.3	66.7
	data warehouse	1	8.3	8.3	75.0
	None	2	16.7	16.7	91.7
	Others	1	8.3	8.3	100.0
	Total	12	100.0	100.0	

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

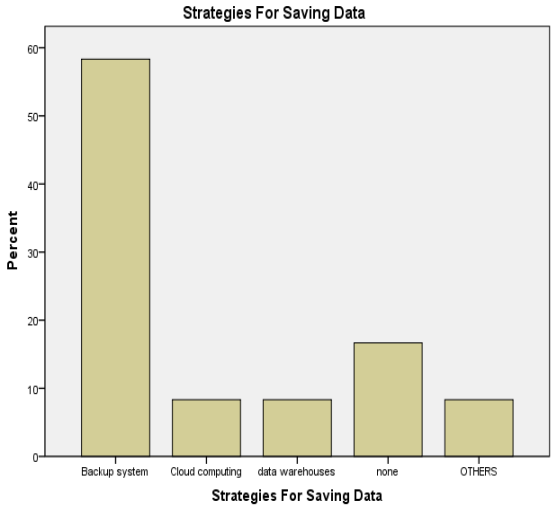


Figure8. Illustrate that strategy for saving data

C. *Others Strategies For Saving Data*

This table shows that others strategies for saving data. Almost 8.3% organizations of Pakistan using cloud server and Oracle to save their data. The results also have been shown in Table 4 and in Figure 9.

TABLE IV. Others Strategies for saving data

Others Strategies for Saving Data

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		10	83.3	83.3	83.3
	cloud server	1	8.3	8.3	91.7
	Oracle	1	8.3	8.3	100.0
	Total	12	100.0	100.0	

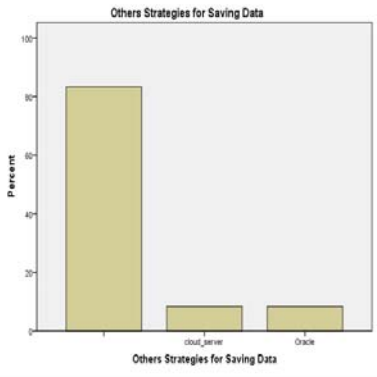


Figure9. Illustrate those others strategies for saving data

D. *Big Data Tools For Saving Data*

In this table shows that only 16.7% organizations of Pakistan using Hadoop, 8.3 using Jaspersoft and 8.3% using Talend Open Studio and 66.7% organizations not using big data tools for saving data. They are all using others strategies for saving their data. The results also have been shown in Table 5 and in Figure 10.

TABLE V. Big Data Tools For Saving Data

Big Data Tools

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Hadoop	2	16.7	16.7	16.7
	Jaspersoft BI Suite	1	8.3	8.3	25.0
	Talend Open Studio	1	8.3	8.3	33.3
	Others	8	66.7	66.7	100.0
	Total	12	100.0	100.0	

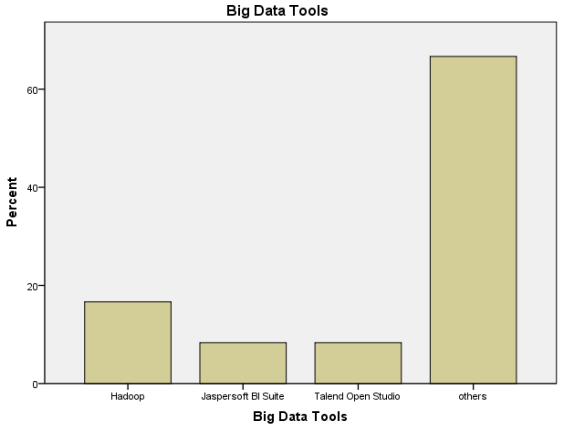


Figure10. Illustrate that big data tools for saving data

E. *Using others tools rather than Big* **Data**

In this table shows that 25.0% organizations of Pakistan using cloud computing, 25% using Oracle and 16.7% using SQL for saving their data. The results also have been shown in Table 6 and in Figure 11.

TABLE VI. Others Tools Rather Than Big Data

Using Others tools rather than Big Data

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		4	33.3	33.3	33.3
	Cloud based	1	8.3	8.3	41.7

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

	Cloud Computing	2	16.7	16.7	58.3
	Oracle	3	25.0	25.0	83.3
	SQL	2	16.7	16.7	100.0
	Total	12	100.0	100.0	

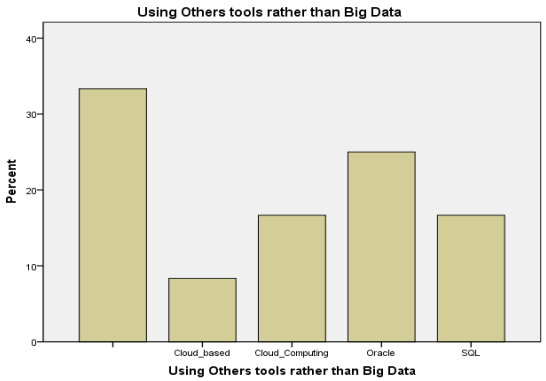


Figure11. Illustrate that organizations using others tools rather than big data

F. Generating Source for data recording

This table shows that 25% organizations of Pakistan using desktop as a generating source for data recording and 75% organizations are using both desktop and laptop for data recording. The results also have been shown in Table 7 and in Figure 12.

TABLE VII. Generating Source For Data Recording

Generating Source for Data Recording

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	desktop	3	25.0	25.0	25.0
	both	9	75.0	75.0	100.0
	Total	12	100.0	100.0	

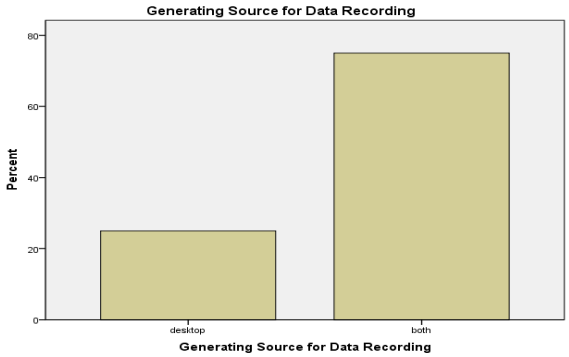


Figure12. Illustrate that data recording sources

G. Format for Information Extracting and Changing

This table shows methods for information extracting and changing, 16.7% organizations of Pakistan using archieve file, 8.3% using default compression, 33.3% using data extractions tools and 41.7% using others strategies for information extracting and changing. The results also have been shown in Table 8 and in Figure 13.

TABLE VIII. Information Extracting And Changing

Information Extracting and Changing

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Archieve file	2	16.7	16.7	16.7
	default compression	1	8.3	8.3	25.0
	data extraction tools	4	33.3	33.3	58.3
	Others	5	41.7	41.7	100.0
	Total	12	100.0	100.0	

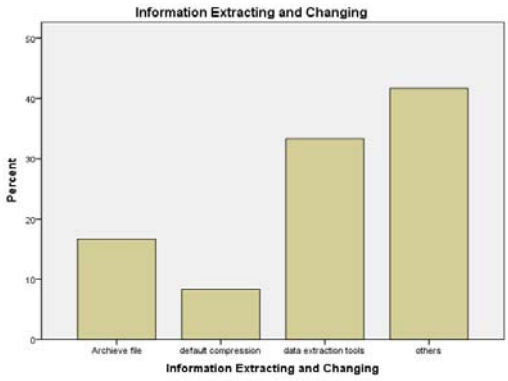


Figure13. Illustrate information extracting and changing

Figure 14 has been shown about power source for system running always. In this figure #1 shows “Generator Power Source”, #2 shows “UPS Power Source”, #3 shows

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

"Both" and #4 shows "none". During my survey, I found the result only one educational institution using one source and all others using both source for always running systems. The result has been shown in Figure 14.

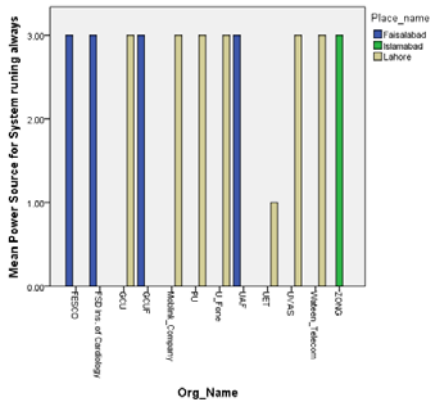


Figure14. Illustrate power source for system running

Figure 15 has been shown about the Database Application System. In this figure #1 shows "web based", #2 shows "desktop", #3 shows "manual" and #4 shows "all of these". During my survey, I found the following result. The result has been shown in Figure 15.

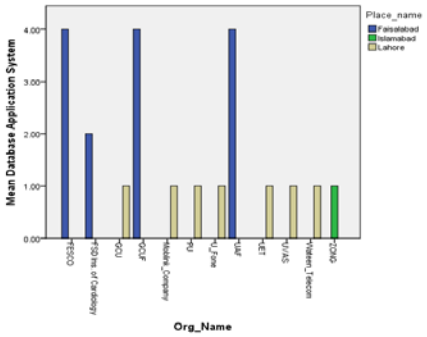


Figure15. Illustrate Database Applications System

Figure 16 has been shown about the Database System. In this figure #1 shows "SQL", #2 shows "Oracle", #3 shows "IBM Data Warehouse" and #4 shows "others". During my survey, I found the following result. The result has been shown in Figure 16.

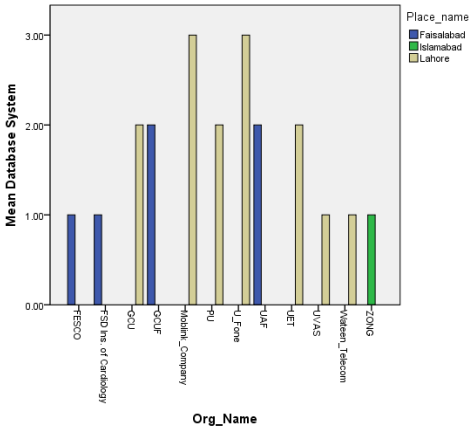


Figure16. Illustrate Database system

Chi test associations between client servers based system and database locking

Research Hypothesis (HI) Client-Servers based system and database locking system relate to each other

Significance level=0.05%

Client-Server Based System * Database Locking System Cross tabulation

			Database Locking System				Total
			pessimistic	optimistic	both	none	
Client-Server Based System	yes	Count	3	3	4	0	10
		Expected Count	2.5	2.5	3.3	1.7	10.0
		% within Client-Server Based System	30.0%	30.0%	40.0%	.0%	100.0%
	no	Count	0	0	0	2	2
		Expected Count	.5	.5	.7	.3	2.0
		% within Client-Server Based System	.0%	.0%	.0%	100.0%	100.0%
Total		Count	3	3	4	2	12
		Expected Count	3.0	3.0	4.0	2.0	12.0
		% within Client-Server Based System	25.0%	25.0%	33.3%	16.7%	100.0%

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Client-Server Based System * Database Locking System	12	100.0%	0	.0%	12	100.0%

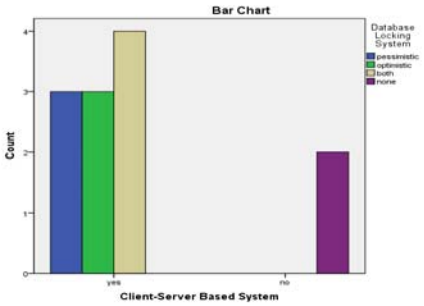


Figure17. Client-Server Based System

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	12.000[a]	3	.007
Likelihood Ratio	10.813	3	.013
Linear-by-Linear Association	5.124	1	.024
N of Valid Cases	12		

a. 8 cells (100.0%) have expected count less than 5. The minimum expected count is .33.

*Chi test associations between DBA in Organizations * Experience of DBA*

Research Hypothesis (HI) DBA in Organizations and Experience of DBA

Significance level=0.05%

DBA In Organizations * Experience of DBA Cross tabulation

			Experience of DBA			Total
			1-5	6-10	11-15	
DBA In Organizations	1	Count	1	6	0	7
		Expected Count	1.8	4.7	.6	7.0
		% within DBA In Organizations	14.3%	85.7%	.0%	100.0%
	2	Count	2	1	0	3
		Expected Count	.8	2.0	.3	3.0
		% within DBA In Organizations	66.7%	33.3%	.0%	100.0%
	3	Count	0	1	1	2
		Expected Count	.5	1.3	.2	2.0
		% within DBA In Organizations	.0%	50.0%	50.0%	100.0%
Total		Count	3	8	1	12
		Expected Count	3.0	8.0	1.0	12.0
		% within DBA In Organizations	25.0%	66.7%	8.3%	100.0%

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
DBA In Organizations * Experience of DBA	12	100.0%	0	.0%	12	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	8.869[a]	4	.064
Likelihood Ratio	7.442	4	.114
Linear-by-Linear Association	.590	1	.442
N of Valid Cases	12		

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	8.869[a]	4	.064
Likelihood Ratio	7.442	4	.114
Linear-by-Linear Association	.590	1	.442
N of Valid Cases	12		

a. 9 cells (100.0%) have expected count less than 5. The minimum expected count is .17.

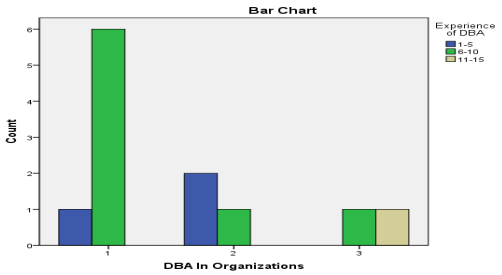


Figure18. DBA in Organizations

*Chi test associations between Data Recovery Method after data Lost * Level of Backup*

Research Hypothesis (HI) Data Recovery Method after data Lost and Level of Backup Cross Significance level=0.05%

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Data Recovery Method after data Lost * Level of Backup	12	100.0%	0	.0%	12	100.0%

Data Recovery Method after data Lost * Level of Backup Cross tabulation

			Level of Backup			Total
			full backup	offline backup	online backup	
Data Recovery Method after data Lost	backup	Count	5	1	1	7
		Expected Count	5.8	.6	.6	7.0
		% within Data Recovery Method after data Lost	71.4%	14.3%	14.3%	100.0%
	recovery method	Count	2	0	0	2
		Expected Count	1.7	.2	.2	2.0
		% within Data Recovery Method after data Lost	100.0%	.0%	.0%	100.0%
	both	Count	3	0	0	3
		Expected Count	2.5	.3	.3	3.0
		% within Data Recovery Method after data Lost	100.0%	.0%	.0%	100.0%
Total		Count	10	1	1	12
		Expected Count	10.0	1.0	1.0	12.0
		% within Data Recovery Method after data Lost	83.3%	8.3%	8.3%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.714[a]	4	.788
Likelihood Ratio	2.438	4	.656
Linear-by-Linear Association	1.041	1	.307
N of Valid Cases	12		

a. 8 cells (88.9%) have expected count less than 5. The minimum expected count is .17.

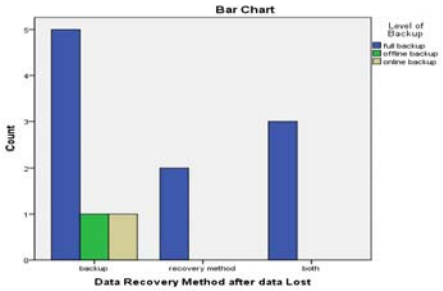


Figure19. Data recovery method after data lost

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

V. Conclusion

Big data related problems are faced by almost all organizations in the world. But I was survey only twelve organizations in Pakistan and create a result using SPSS software. In my report almost results are significant and Ho is rejected.

Many organizations still used olds methods. Some organizations have no knowledge about big data. In Pakistan 75% organizations saved their data using computer system. But yet have no idea about big data usage. In Pakistan 16.7% organizations using Hadoop, 8.3 using Jaspersoft and 8.3% using Talend Open Studio and 66.7% organizations still not using big data tools for saving data. They are all using others strategies for saving their data. 25% organizations using desktop as a generating source for data recording and 75% organizations are using both desktop and laptop for data recording. 16.7% organizations using archieve file, 8.3% using default compression, 33.3% using data extractions tools and 41.7% using others strategies for information extracting and changing. Methods percentage for data cleaning which is used by organizations in Pakistan. 25% organizations using data mining tools, 16.7% using batch processing, 33.3% using others tools and 25% using no tools for data cleaning. 50% organizations using SQL query, 41.7% are using both SQL and PLSQL and 8.3% using no methods for querying data. 25% organizations using WEKA tools and 75% using others tools for mining their data.

ACKNOWLEDGEMENTS

The support provided by the Department of Information Technology, Govt. College University Faisalabad, Pakistan, is appreciatively acknowledged. Authors would also like to express thanks my respected teacher, for her precious suggestions and criticisms. In addition, authors would like to thanks my entire family members, who support me in my study and research.

REFERENCES

[1] S. Kaisler, F. Armour, J. Alberto Espinosa, W. Money, (2013). 46th Hawaii International Conference on System Sciences. “Big Data: Issues and Challenges Moving Forward”, Journal of IEEE, pp.645.

[2] Hashem, Ibrahim Abaker Targio, et al (2015). "The rise of “big data” on cloud computing: Review and open research issues." *Information Systems* 47 pp.98-115.

[3] Mervis, J. (2012). “Agencies Rally to Tackle Big Data”, Science, pp.336 (4):22, June 6.

[4] Oracle and FSN, (December 2012). "Mastering Big Data: CFO Strategies to Transform Insight into Opportunity".

[5] Rubinstein, Ira S., (2012). "Big Data: The End of Privacy or a New Beginning?" New York University Public Law and Legal Theory Working Papers.Paper pp.357.

[6] Reips, Ulf-Dietrich; Matzat, Uwe *Mining (*2014). *"Big Data" using Big Data Services”. International Journal of Internet Science.**1** (1):pp. 1–8.*

[7] Torgerson, P R; Torgerson, D (2010). Public health and bovine tuberculosis: what's all the fuss about? Trends in Microbiology, 18(2):pp.67-72.

[8] Alexandru Adrian TOLE, (2013). Romanian American University, Bucharest, Romania, “Big Data Challenges “*Database Systems Journal* vol. IV, no. 3.

[9] Katal. A, Wazid. M and R H Gouder, (2013). ”Big Data: Issues, Challenges, Tools and Good Practices”.pp.978- IEEE.

[10] Suthaharan, S. (2014). “Big data classification: Problems and challenges in network intrusion prediction with machine learning”. Performance Evaluation Review, 41(4), pp.70-73.

[11] Tilmann. R, Sadoghi. M and Hans-Aron.J, (21 Aug 2012). “Solving Big Data Challenges for Enterprise Application Performance Management.ar”.

[12] Rubinstein, Ira S., (2012). "Big Data: The End of Privacy or a New Beginning?" .New York University Public Law and Legal Theory Working pp. 357.

[13] Boyd, Dana; Crawford, Kate, (September 21, 2011). "Six Provocations for Big Data". *Social Science Research Network: A Decade in Internet Time: Symposium on the Dynamics of the InternetandSociety*.

[14] Daniel, Ben, (2015). "Big data and analytics in higher education: Opportunities and challenges." *British journal of educational technology* 46.5 pp.904-920.

[15] Demchenko.Y and Membrey. P, (2013) “Defining Architecture Components of the Big Data Ecosystem”. System and Network Engineering Group University of Amsterdam, The Netherlands e-mail: {y.demchenko, C.T.A.M.deLaat}@uva.nl, Hong Kong Polytechnic University Hong Kong SAR, China e-mail: cspmembrey@comp.polyu.edu.h.

[16] Y. Chen, S. Alspaugh, R. Katz, (August 27th - 31st 2012). “Interactive Analytical Processing in Big Data Systems: A Cross-Industry Study of Map Reduce Workloads”, Istanbul, Turkey. Proceedings of the VLDB Endowment, Vol. 5, pp. 12.

[17] S.S Sahoo, C. Jayapandian, G. Garg, F. Kaffashi, S. Chung, A. Bozorgi, (2013). "Heart beats in the cloud: distributed analysis of electrophysiological ‘Big Data ‘using cloud computing for epilepsy clinical research." *Journal of the American Medical Informatics Association* 21.2: pp.263-271.

[18] Raghupathi, Wullianallur, and Viju Raghupathi, (2014). "Big data analytics in healthcare: promise and potential." *Health information science and systems pp.*2.1, (2014): 3.

[19] Liu, Yao, Peng Ning, and Michael K. Reiter, (2011). "False data injection attacks against state estimation in electric power grids." *ACM Transactions on Information and System Security (TISSEC)* pp.14.1, (2011): 13.

[20] Lee, Jay; Lapira, Edzel; Bagheri, Behrad; Kao, Hung-An, (2013). *"Recent Advances and Trends in Predictive Manufacturing Systems in Big Data Environment"*. Manufacturing Letters. ***1** (1): pp.38–41.*

[21] Magoulas, Roger; Lorica, Ben, (February 2009). "Introduction to Big Data". *Release 2.0*. Sebastopol CA: O'Reilly Media (11).

[22] Andrea De Mauro, Marco Greco and Michele Grimaldi, (2015). “What is Big Data? A Consensual Definition and a Review of Key Research Topics”, AIP Conference Proceedings 1644, 97.

[23] Al-Rodhan, Nayef (2014-09-16). "The Social Contract 2.0: Big Data and the Need to Guarantee Privacy and Civil Liberties Harvard International Review". *Harvard International Review*. Retrieved pp.04-03.

[24] Gandomi, Amir, and Murtaza Haider, (2015). "Beyond the hype: Big data concepts, methods, and analytics." *International Journal of Information Management*35.2:pp.137-144.

[25] Haiping Lu, K. N. Plataniotis and A. N. Venetsanopoulos, (Jul. 2011). “A Survey of Multilinear Subspace Learning for Tensor Data”, Pattern Recognition, vol. 44, no. 7, pp. 1540-1551.

[26] Hsinchun Chen, Roger H. L. Chiang, Veda C. Storey, (December 2012). “Business Intelligence and Analytics: From Big Data to Big Impact”. *MIS Quarterly Vol. 36 No. 4, pp. 1165-1188.*

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

[27] Nunan, Daniel and Di Domenico, (2015). “M. Market research the ethics of big data”. International Journal of Market Research 55 (4), pp. 505-520. ISSN 1470-7853.